CSC 261/461
Database Systems

Eustrat Zhupa

November 7, 2018

Relational Algebra

CARTESIAN PRODUCT Operation

- This operation is used to combine tuples from two relations.
- Denoted by $R(A1, A2, ..., An)$ x $S(B1, B2, ..., Bm)$
- Result is a relation Q with $n + m$ attributes:
  Q(A1, A2, . . ., An, B1, B2, . . ., Bm)
- The resulting relation state has one tuple for each combination of tuples.
- The two operands do NOT have to be "type compatible"

[1]

Example

Example

EMPNAMES

Fname	Lname	Ssn
Alicia	Zelaya	999887777
Jennifer	Wallace	987654321
Joyce	English	453453453

DEPENDENT

Essn	Dependent_name	Sex	Bdate	Relationship
333445555	Alice	F	1986-04-05	Daughter
333445555	Theodore	M	1983-10-25	Son
333445555	Joy	F	1958-05-03	Spouse
987654321	Abner	M	1942-02-28	Spouse
123456789	Michael	M	1988-01-04	Son
123456789	Alice	F	1988-12-30	Daughter
123456789	Elizabeth	F	1967-05-05	Spouse

Example

Example

EMP_DEPENDENTS

Fname	Lname	Ssn	Essn	Dependent_name	Sex	Bdate	...
Alicia	Zelaya	999887777	333445555	Alice	F	1986-04-05	...
Alicia	Zelaya	999887777	333445555	Theodore	M	1983-10-25	...
Alicia	Zelaya	999887777	333445555	Joy	F	1958-05-03	...
Alicia	Zelaya	999887777	987654321	Abner	M	1942-02-28	...
Alicia	Zelaya	999887777	123456789	Michael	M	1988-01-04	...
Alicia	Zelaya	999887777	123456789	Alice	F	1988-12-30	...
Alicia	Zelaya	999887777	123456789	Elizabeth	F	1967-05-05	...
Jennifer	Wallace	987654321	333445555	Alice	F	1986-04-05	...
Jennifer	Wallace	987654321	333445555	Theodore	M	1983-10-25	...
Jennifer	Wallace	987654321	333445555	Joy	F	1958-05-03	...
Jennifer	Wallace	987654321	987654321	Abner	M	1942-02-28	...
Jennifer	Wallace	987654321	123456789	Michael	M	1988-01-04	...
Jennifer	Wallace	987654321	123456789	Alice	F	1988-12-30	...
Jennifer	Wallace	987654321	123456789	Elizabeth	F	1967-05-05	...
Joyce	English	453453453	333445555	Alice	F	1986-04-05	...
Joyce	English	453453453	333445555	Theodore	M	1983-10-25	...
Joyce	English	453453453	333445555	Joy	F	1958-05-03	...
Joyce	English	453453453	987654321	Abner	M	1942-02-28	...
Joyce	English	453453453	123456789	Michael	M	1988-01-04	...
Joyce	English	453453453	123456789	Alice	F	1988-12-30	...
Joyce	English	453453453	123456789	Elizabeth	F	1967-05-05	...

Operations of RA

JOIN

- Denoted by $\bowtie$
- The sequence of CROSS PRODUCT followed by SELECT is used quite to identify and select related tuples from two relations
- A special operation, called JOIN combines them into a *single* operation
- very important because it allows us combine data from various relations
- The general form of a join operation is:

$$R \bowtie_{<joincondition>} S$$

Operations of RA

JOIN

- Consider the following JOIN:
  $R(A1, A2, ..., An) \bowtie_{R.Ai=S.Bj} S(B1, B2, ..., Bm)$
- Result is a relation Q with n + m attributes: Q(A1, A2, . . ., An, B1, B2, . . ., Bm), in that order.
- The result has one tuple for each combination of tuples - one from R and one from S, but only if they satisfy the join condition $r[Ai] = s[Bj]$

[2]

Example

Example

EMPLOYEE

Fname	Minit	Lname	Ssn	Bdate	Address	Sex	Salary	Super_ssn	Dno
John	B	Smith	123456789	1965-01-09	731 Fondren, Houston, TX	M	30000	333445555	5
Franklin	T	Wong	333445555	1955-12-08	638 Voss, Houston, TX	M	40000	888665555	5
Alicia	J	Zelaya	999887777	1968-01-19	3321 Castle, Spring, TX	F	25000	987654321	4
Jennifer	S	Wallace	987654321	1941-06-20	291 Berry, Bellaire, TX	F	43000	888665555	4
Ramesh	K	Narayan	666884444	1962-09-15	975 Fire Oak, Humble, TX	M	38000	333445555	5
Joyce	A	English	453453453	1972-07-31	5631 Rice, Houston, TX	F	25000	333445555	5
Ahmad	V	Jabbar	987987987	1969-03-29	980 Dallas, Houston, TX	M	25000	987654321	4
James	E	Borg	888665555	1937-11-10	450 Stone, Houston, TX	M	55000	NULL	1

DEPARTMENT

Dname	Dnumber	Mgr_ssn	Mgr_start_date
Research	5	333445555	1988-05-22
Administration	4	987654321	1995-01-01
Headquarters	1	888665555	1981-06-19

DEPT_MGR

Dname	Dnumber	Mgr_ssn	...	Fname	Minit	Lname	Ssn	...
Research	5	333445555	...	Franklin	T	Wong	333445555	...
Administration	4	987654321	...	Jennifer	S	Wallace	987654321	...
Headquarters	1	888665555	...	James	E	Borg	888665555	...

Operations of RA

THETA JOIN

- The general case of JOIN operation is called a Theta-join:
  $$R \bowtie_{theta} S$$
- The join condition is called `theta`
- Theta can be any boolean expression on attributes of R and S:
  Example:
  $$R.A_i < S.B_j \ AND \ (R.A_k = S.B_i \ OR \ R.A_p < S.B_q)$$

Example

EQUIJOIN

- Join conditions with equality comparisons only
- In the result we always have one or more pairs of attributes that have identical values in every tuple.

$$DEPT\_MGR \leftarrow DEPARTMENT \bowtie_{Mgr\_ssn=Ssn} EMPLOYEE$$

DEPT_MGR

Dname	Dnumber	Mgr_ssn	...	Fname	Minit	Lname	Ssn	...
Research	5	333445555	...	Franklin	T	Wong	333445555	...
Administration	4	987654321	...	Jennifer	S	Wallace	987654321	...
Headquarters	1	888665555	...	James	E	Borg	888665555	...

Operations of RA

NATURAL JOIN

- denoted by *
- created to get rid of the second attribute in an EQUIJOIN condition.
  - one of each pair of attributes with identical values is redundant
- each pair of corresponding join attributes have the same name in both relations
  - If not, a renaming operation is applied first.

[3]

Example

Example

PROJECT

Pname	Pnumber	Plocation	Dnum
ProductX	1	Bellaire	5
ProductY	2	Sugarland	5
ProductZ	3	Houston	5
Computerization	10	Stafford	4
Reorganization	20	Houston	1
Newbenefits	30	Stafford	4

DEPARTMENT

Dname	Dnumber	Mgr_ssn	Mgr_start_date
Research	5	333445555	1988-05-22
Administration	4	987654321	1995-01-01
Headquarters	1	888665555	1981-06-19

$$PROJ\_DEPT \leftarrow PROJECT * \rho_{(Dname, Dnum, Mgr\_ssn, Mgr\_start\_date)}(DEPARTMENT)$$

(a)

PROJ_DEPT

Pname	Pnumber	Plocation	Dnum	Dname	Mgr_ssn	Mgr_start_date
ProductX	1	Bellaire	5	Research	333445555	1988-05-22
ProductY	2	Sugarland	5	Research	333445555	1988-05-22
ProductZ	3	Houston	5	Research	333445555	1988-05-22
Computerization	10	Stafford	4	Administration	987654321	1995-01-01
Reorganization	20	Houston	1	Headquarters	888665555	1981-06-19
Newbenefits	30	Stafford	4	Administration	987654321	1995-01-01

Example

Example

DEPARTMENT

Dname	Dnumber	Mgr_ssn	Mgr_start_date
Research	5	333445555	1988-05-22
Administration	4	987654321	1995-01-01
Headquarters	1	888665555	1981-06-19

DEPT_LOCATIONS

Dnumber	Dlocation
1	Houston
4	Stafford
5	Bellaire
5	Sugarland
5	Houston

DEPT_LOCS $\leftarrow$ *DEPARTMENT* $*$ *DEPT_LOCATIONS*?

(b)

DEPT_LOCS

Dname	Dnumber	Mgr_ssn	Mgr_start_date	Location
Headquarters	1	888665555	1981-06-19	Houston
Administration	4	987654321	1995-01-01	Stafford
Research	5	333445555	1988-05-22	Bellaire
Research	5	333445555	1988-05-22	Sugarland
Research	5	333445555	1988-05-22	Houston

Operations of RA

OUTER JOIN

- LEFT OUTER JOIN $R \ltimes\!\!\!\rightarrow S$: keeps every tuple in R; if no matching tuple is found in S, then the attributes of S are filled with NULL.
- RIGHT OUTER JOIN $R \ltimes\!\!\!\leftarrow S$: keeps every tuple in S in the result, fill in with NULLs in R.
- FULL OUTER JOIN $R \ltimes\!\!\!\rightarrow\!\!\!\leftarrow S$: keeps all tuples in both relations: when no matching tuples are found, fills with NULL.

[4]

Example

Example

EMPLOYEE

Fname	Minit	Lname	Ssn	Bdate	Address	Sex	Salary	Super_ssn	Dno
John	B	Smith	123456789	1965-01-09	731 Fondren, Houston, TX	M	30000	333445555	5
Franklin	T	Wong	333445555	1955-12-08	638 Voss, Houston, TX	M	40000	888665555	5
Alicia	J	Zelaya	999887777	1968-01-19	3321 Castle, Spring, TX	F	25000	987654321	4
Jennifer	S	Wallace	987654321	1941-06-20	291 Berry, Bellaire, TX	F	43000	888665555	4
Ramesh	K	Narayan	666884444	1962-09-15	975 Fire Oak, Humble, TX	M	38000	333445555	5
Joyce	A	English	453453453	1972-07-31	5631 Rice, Houston, TX	F	25000	333445555	5
Ahmad	V	Jabbar	987987987	1969-03-29	980 Dallas, Houston, TX	M	25000	987654321	4
James	E	Borg	888665555	1937-11-10	450 Stone, Houston, TX	M	55000	NULL	1

DEPARTMENT

Dname	Dnumber	Mgr_ssn	Mgr_start_date
Research	5	333445555	1988-05-22
Administration	4	987654321	1995-01-01
Headquarters	1	888665555	1981-06-19

RESULT

Fname	Minit	Lname	Dname
John	B	Smith	NULL
Franklin	T	Wong	Research
Alicia	J	Zelaya	NULL
Jennifer	S	Wallace	Administration
Ramesh	K	Narayan	NULL
Joyce	A	English	NULL
Ahmad	V	Jabbar	NULL
James	E	Borg	Headquarters

Division

Division

The division operator denoted as $R \div S$.

- If $R(Z) \div S(X)$, where $X \subseteq Z$, and $Y = Z - X$.
- Result is a relation $T(Y)$ including a tuple t if:
  - t_R appears in R with $t_R[Y] = t$
  - $t_R[x] = t_S$ for every tuple t_S in S.
- Otherwise, For a tuple t to appear in the result, values in t must be in R combined with all tuples in S.

Division

Example

SSN_PNOS

Essn	Pno
123456789	1
123456789	2
666884444	3
453453453	1
453453453	2
333445555	2
333445555	3
333445555	10
333445555	20
999887777	30
999887777	10
987987987	10
987987987	30
987654321	30
987654321	20
888665555	20

SMITH_PNOS

Pno
1
2

SSNS

Ssn
123456789
453453453

R

A	B
a1	b1
a2	b1
a3	b1
a4	b1
a1	b2
a3	b2
a2	b3
a3	b3
a4	b3
a1	b4
a2	b4
a3	b4

S

A
a1
a2
a3

T

B
b1
b4

RA Operations

Generalized Projection

The generalized projection operation:

$$\pi_{F_1, F_2, \ldots, F_n}(R)$$

- $F_1, F_2, \ldots, F_n$ are functions over the attributes of R and may involve arithmetic operations and constant values.

$$\pi_{Ssn, Salary - Deduction, 2000 * Years\_service, 0.25 * Salary}(EMPLOYEE)$$

RA Operations

Aggregate Functions

We can define an AGGREGATE FUNCTION operation, using the symbol $\mathcal{F}$:

$$<grouping\_attributes>\mathcal{F}_{<function\_list>}(R)$$

- $<$ *grouping_attributes* $>$ is a list of attributes
- $<$ *function_list* $>$ is a list of ($<$ *function* $>$, $<$ *attribute* $>$) pairs.
- $<$ *function* $>$ is one of SUM , AVERAGE , MAXIMUM , MINIMUM , COUNT.

Example

R

(a)

Dno	No_of_employees	Average_sal
5	4	33250
4	3	31000
1	1	55000

(b)

Dno	Count_ssn	Average_salary
5	4	33250
4	3	31000
1	1	55000

(c)

Count_ssn	Average_salary
8	35125

Figure 8.10
The aggregate function operation.

a. $\rho_{R(\text{Dno, No\_of\_employees, Average\_sal})}(_{\text{Dno}}\Im_{\text{COUNT Ssn, AVERAGE Salary}}(\text{EMPLOYEE}))$.

b. $_{\text{Dno}}\Im_{\text{COUNT Ssn, AVERAGE Salary}}(\text{EMPLOYEE})$.

c. $\Im_{\text{COUNT Ssn, AVERAGE Salary}}(\text{EMPLOYEE})$.

UNIVERSITY of ROCHESTER

?